

Empirical Analysis and Evolving Model of Bipartite Networks

Peng Zhang^{1,2}, Menghui Li¹, J.F.F. Mendes², Zengru Di¹, Ying Fan^{1*},

1. Department of Systems Science, School of Management,
Beijing Normal University, Beijing 100875, P.R.China
2. Departamento de Física da Universidade de Aveiro,
3810-193 Aveiro, Portugal

April 24, 2008

Abstract

Many real-world networks display a natural bipartite structure. Investigating it based on the original structure is helpful to get deep understanding about the networks. In this paper, some real-world bipartite networks are collected and divided into two types, dependence bipartite networks and independence bipartite networks, according to the different relation of two sets of nodes. By analyzing them, the results show that the actors nodes have scale-free property in the dependence networks, and there is no accordant degree distribution in the independence networks for both two types of nodes. In order to understand the scale-free property of actors in dependence networks, two growing bipartite models without the preferential attachment principle are proposed. The models show the scale-free phenomena in actors' degree distribution. It also gives well qualitatively consistent behavior with the empirical results.

Keyword: Bipartite networks, Evolving model,

1 Introduction

In recent years, the complex networks has attracted more and more people's attention [1, 2, 3, 4]. Many real-world systems are depicted as complex networks to investigate their structures and functions. Examples include WWW, internet, food webs, biochemical networks, social networks, and so on [5, 6, 7, 8, 9, 10, 11]. These researches in networks not only raised new concepts and methods, but also helped us understand complex systems.

*Author for correspondence: yfan@bnu.edu.cn

Bipartite network is an important kind of complex networks. In fact, many real-world networks display a natural bipartite structure, Examples are the actors-films network [12], the papers-scientists network [13, 14, 15] and so on. A bipartite graph is defined as a network in which nodes are divided into two sets, one representing a set of actors, while the other representing acts of collaboration, so that no edge connects two nodes in the same set. In other words, The edges only connect a pair of nodes belongs to different sets. These networks are also called 2-mode networks, or affiliation networks when they represent groups and members (*i.e.* each link represents a social actors affiliation to a group) [16, 17]. In the present paper, following the relation of two sets of nodes, we consider that bipartite networks can be separated into two types of networks, dependence bipartite networks and independence bipartite networks. In dependence networks, two sets are dependent. The groups are the aftermaths of the behaviors among actors, the actors lead to the groups. Examples of the dependency networks are the actors-films network [12], where a film is a production of the cooperation among actors, the papers-scientists network [13, 14, 15], where some scientists coauthor a paper, and the occurrence networks, where a sentence is organized by some words. The independence network is opposite with the dependence network, two sets exist independently. Let us cite for instance the booker-reader network, where the books always are as the resource in library whatever the readers do, the sex networks [18], where the male and the female do not affect each other, and the peer to peer networks [19, 20], where as the downloader and the uploader, they are not the output of each other.

Now many real world networks have been analyzed as one mode networks, and been found that have the scale-free property [14, 15, 21]. However most of them has not been done on the generation of complex networks with the scale-free property expressed by bipartite graphs. Newman *et al* [22]. have proposed a model for generating a complex bipartite graph, in which two degree distributions for both kinds of nodes are given and edges are randomly connected under the constraint that their degree distributions are fixed. Ramasco *et al*.[21] Guillaume *et al*[23, 24]. Ramezanpour [25]. Goldstein *et al*[13]. and Lambiotte *et al*[26] have proposed the other type of models for generation of complex bipartite graphs; those are growing networks in which the number of nodes and that of edges are increasing and newly added edges are connected to nodes by a suitable rule. In order to generate a complex network of bipartite graph with the scale-free property, Goldstein *et al*. Ramasco *et al*. and Guillaume *et al*. have applied a rule, which is based on the concept of the preferential attachment proposed by Barabási and Albert [27], to their model for generating complex networks. There also have the other type of bipartite networks in which the size of networks is fixed, but the structure of networks changes rapidly. Ohkubo *et al* [28]. note this and proposed a model to generate complex bipartite graphs without growing. The essential ingredients are a preferential rewiring process and a fitness distribution function in this model. They showed that the model generates the complex networks with the scale-free property for some fitness distribution functions.

In the present paper, we collect different types of real world data and build them

to complex bipartite networks. By analyzing these bipartite networks, we try to find the common properties in the dependency networks (in the independence networks). Because in bipartite networks, there is a lack of standard definitions of some properties that we often analyze in one-mode networks, here we only focus on the degree distribution of both types of nodes. The results reveal that the actors nodes have scale-free property in the dependence networks, and there is no accordant degree distribution in the independence networks for both two types of nodes. Then we propose two growing models of complex bipartite networks for the dependence networks without preferential attachment. The degree distribution of actors in the networks obtained with both models can capture the scale-free property which appears in real dependence networks.

The outline of this paper is as follow. In Section 2, we introduce two types of networks in complex bipartite networks, the dependence networks and the independence networks. Some real world networks of them are built and analyzed. The results show that the actors nodes have scale-free property in the dependence networks. We propose two growing models of complex bipartite networks for the dependence networks in Section 3. It also contains the simulation results. We obtain the degree distribution of generated dependence networks, and the scale-free property is found in them. In Section 4, we give our concluding remarks.

2 Bipartite Networks Data

In order to investigate the properties of the dependence networks and the independence networks, here we consider some representation real-world examples of them. The set of the dependence networks includes a Econophysicists bipartite network [29], a complex-networks network, the co-occurrence relation of words in the sentences of the Bible network [30], and a coauthoring relation between scientists network [14, 15, 22, 17]. The independence networks consist of a books-readers database obtained from Beijing Normal University library, a P2P network gotten from *Maze.com*, and another P2P network ¹[19, 20]. To be convenient in next sections, we refer to replace them as *Econophysicists*, *Complex – networks*, *Cooccurrence*, *Coauthoring*, *Library*, *Maze*, and *P2P*. All of them have been defined and studied as the one-mode network well in cited references.

2.1 The Dependence Networks

Econophysicists We collected publish information of Econophysicists papers published from 1992 to 2007, and built the Econophysicists bipartite network with if the author wrote the paper, there would be an edge connects with them. The final bipartite network comprises $N_{authors} = 1963$ authors, $N_{papers} = 2020$ papers and $E = 5129$ edges between them. The average degree of authors and papers are $\langle k_{authors} \rangle =$

¹The choice of two kind nodes are different in these two P2P networks, and the data also gotten from different sites.

$E/N_{authors} = 2.61$ papers per author and $\langle k_{papers} \rangle = E/N_{papers} = 2.54$ authors per paper, while the maximal degree of them respectively are 85 and 8.

Complex-networks The complex networks is a hot topic in recent years. There are many papers about it. We search the papers in SCI web sites with the keyword complex networks from 1998 to 2007 to build this Complex-networks network. Scientists and papers are identified as two different kinds of nodes, an edge exists between a scientist and a paper if the scientist is one of the paper's authors. the resulting bipartite network consists of 50545 scientists, 18833 papers, and 72582 edges. The average degree of them are $\langle k_{scientists} \rangle = 1.436$ and $\langle k_{papers} \rangle = 3.85$, while the maximal degree are 83 and 166.

Cooccurrence The data of this co-occurrence network obtained from a new online version of the Bible. There are two different kinds of nodes, words and sentences, with a word connects to a sentence which it belongs to. The built bipartite network is composed of $N_{words} = 9264$, $N_{sentences} = 13587$, and $E = 183363$. The average degree of words and sentences respectively are $\langle k_{words} \rangle = 19.8$ and $\langle k_{sentences} \rangle = 13.5$, while the maximal degree are 10454 and 52.

Coauthoring This co-authoring bipartite network is defined the same with the complex-networks network. The data obtained from the online *arXiv* preprint repository. It includes $N_{scientists} = 16400$, $N_{papers} = 19885$, and $E = 45904$. The average degree of scientists and papers respectively are $\langle k_{scientists} \rangle = 2.8$ and $\langle k_{papers} \rangle = 2.3$, while the maximal degree are 62 and 8.

The degree distribution of two sets of these four dependence networks are given in Fig. 1–4. Apparently, One may observe on these plots that all these distributions have a property in common: the degree distributions of actors (Shown in Fig. 1(a)–4(a)) are very heterogeneous and fit power laws very well in all cases [17, 21, 23, 24]. On the contrary, the degree distributions of groups (Shown in Fig. 1(b)–4(b)) are far from a power law in these four cases. The reason may be the limit number of authors, words in papers and sentences.

2.2 The Independence Networks

Library Here we analyze a books-readers bipartite network which the database obtained from Beijing Normal University library during 2005, with $N_{readers} = 17593$ and $N_{books} = 91752$. An edge exists between a reader and a book if he/she borrowed this book, and $E = 369394$ in this graph. The average degree of readers and books respectively are $\langle k_{readers} \rangle = 20.99$ and $\langle k_{books} \rangle = 4.026$, while the maximal degree are 410 and 361.

Maze It is a peer to peer exchange bipartite network. In this bipartite network, the data obtained from *Maze.com* by registering all the exchanges processed by a large server during one week. There are two kinds of peers to be two kind nodes. One kind peers are the downloaders who download the resources which provided by the other kind peers. It consists of $N_{downloader} = 110163$, $N_{resources} = 25171$, and $E = 924738$. The

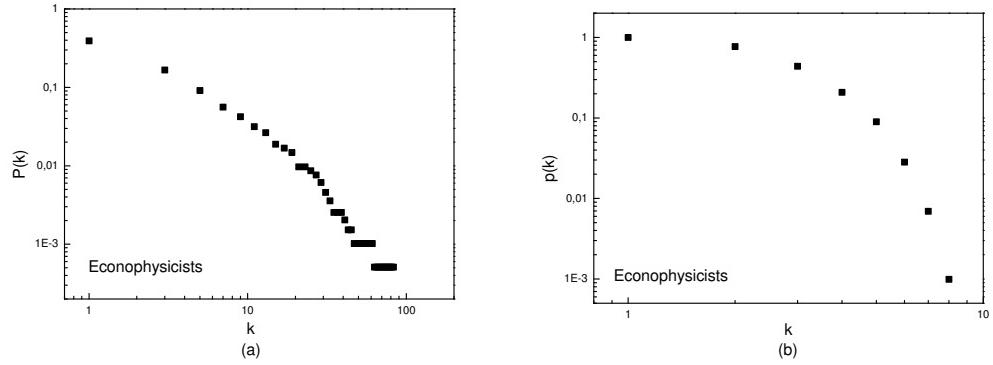


Figure 1: (a) Degree distribution of authors in the Econophysicists network. The degree k corresponds to the number of papers of each author signed. (b) The same distribution of papers in the Econophysicists network. The degree k corresponds to the number of authors of each paper.

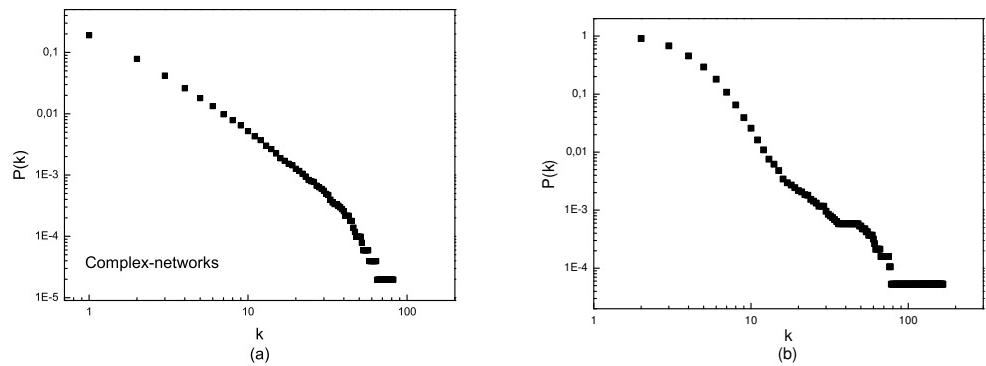


Figure 2: (a) Degree distribution of scientists in the Complex-networks network. The degree k corresponds to the number of papers of each scientist signed. (b) The same distribution of papers in the Complex-networks network. The degree k corresponds to the number of scientists of each paper.

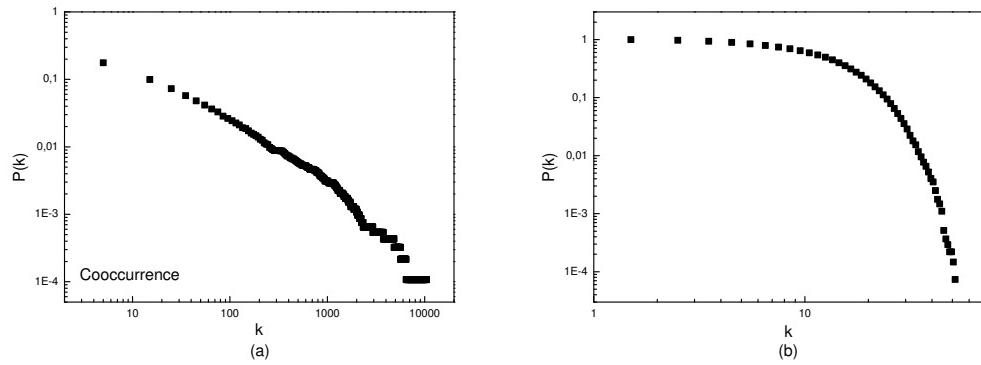


Figure 3: (a) Degree distribution of words in the Cooccurrence network. The degree k corresponds to the number of sentences of each word belonged to. (b) The same distribution of sentences in the Cooccurrence network. The degree k corresponds to the number of words of each sentence.

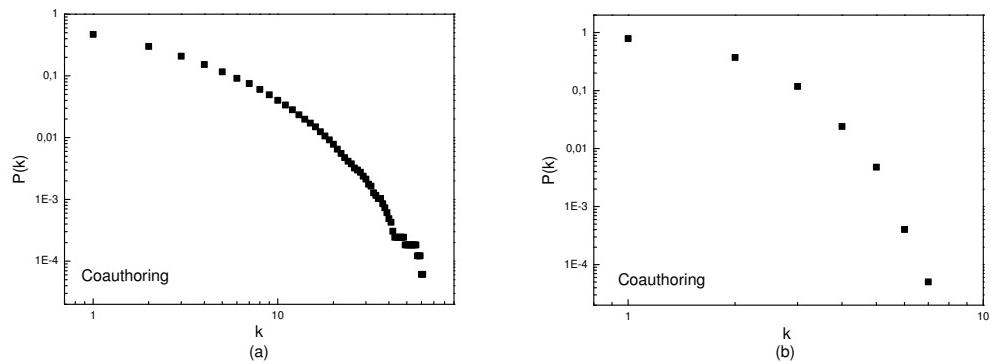


Figure 4: (a) Degree distribution of authors in the Coauthoring network. The degree k corresponds to the number of papers of each author signed. (b) The same distribution of papers in the Coauthoring network. The degree k corresponds to the number of authors of each paper.

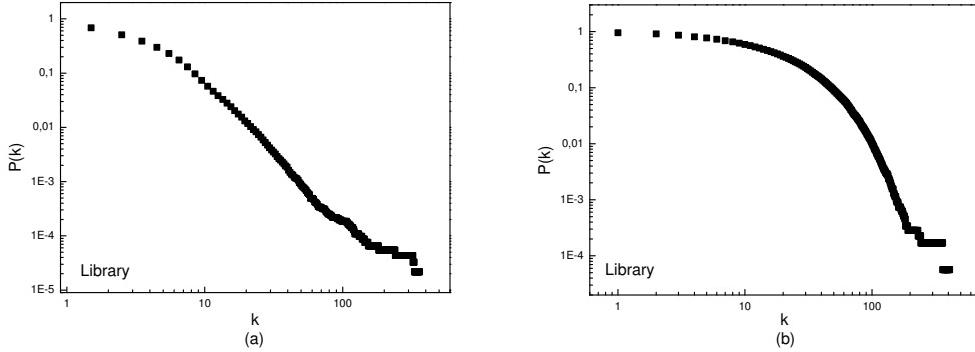


Figure 5: (a) Degree distribution of books in the Library network. The degree k corresponds to the number of readers of each book have been borrowed. (b) The same distribution of readers in the Library network. The degree k corresponds to the number of books of each reader have borrowed.

average degree of downloaders and resources respectively are $\langle k_{\text{downloaders}} \rangle = 8.39$ and $\langle k_{\text{resources}} \rangle = 36.74$, while the maximal degree are 1455 and 1668.

P2P It is also a peer to peer exchange bipartite network. But the differences with *Maze* are not only the data source but also the definition of two kind nodes. Here peers and data are two different kinds of nodes. The data source gotten from all the exchanges processed by a large server during 48 hours, contains $N_{\text{peers}} = 1986588$, $N_{\text{data}} = 5380546$, and $E = 55829392$. The average degree are respectively $\langle k_{\text{peers}} \rangle = 28.1$ and $\langle k_{\text{data}} \rangle = 10.38$.

The degree distribution of two sets of *Library* and *Maze* are given in Fig. 5,6. The real world databases of independence networks are very limit which we could collect. Just from the gotten results in these three cases, it hardly concludes that there are some common properties in the independence networks.

3 The Evolution Model and Simulation Results

3.1 The Evolution Model

In order to understand the scale-free property of actors in dependence networks, most of the former works used the preferential attachment principle to define the models [13, 21, 23, 24, 26]. Here we propose two growing models without the preferential attachment principle to obtain that the authors degree distributions generally follow a power law. To be concrete, we will describe the developments in terms of authors and papers, but our results are applicable to any real affiliation networks and generated networks. The first growing bipartite network model (Model 1) is defined by the following rules:

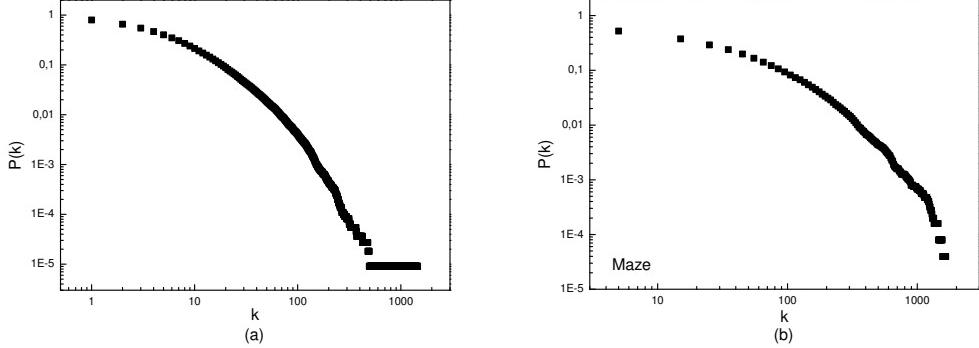


Figure 6: (a) Degree distribution of downloaders in the Maze network. The degree k corresponds to the number of resources of each downloader have downloaded. (b) The same distribution of uploaders in the Maze network. The degree k corresponds to the number of downloaders of each uploader shared.

- (1) At each time step a new paper with n authors is added.
- (2) In these n authors, there is one author is new with probability λ , or not in this time step.
- (3) The rest $n - 1$ or n authors are chosen from the pool of "old" authors. The rule follows: randomly select an old existing author, every other authors is randomly selected from its m th nearest neighborhoods.

The first two steps of the second model (Model 2) are the same as the first one. The big difference between them is the rule of selecting the rest authors.

- (3) The rest $n - 1$ or n authors are chosen from the pool of "old" authors. The rule follows: randomly select an old existing author, every other authors is selected by random walk from it by m steps.

where the parameter λ is a constant. The number n is a random number gotten from 1 to \max_n . The number m is also a random number gotten from 1 to \max_m . \max_n and \max_m are two constants.

3.2 The Simulation Results

In this section, The numerical results of Model 1 and Model 2 are apart given in Figs. 7–10. All the results are the average of 20 simulations for different realization of networks under the same parameters. The step of all generated networks all reach 40000 times. We have also did the simulation to 10000 steps. The final distributions of corresponding quantities are almost same (Shown in Fig. 7). A network with 40000 papers could give us a nice description for asymptotic distribution. Parameters are $\max_n=5$, $\lambda=0.8$, $\max_m=2$ for two models if not mentioned. As the increasing of

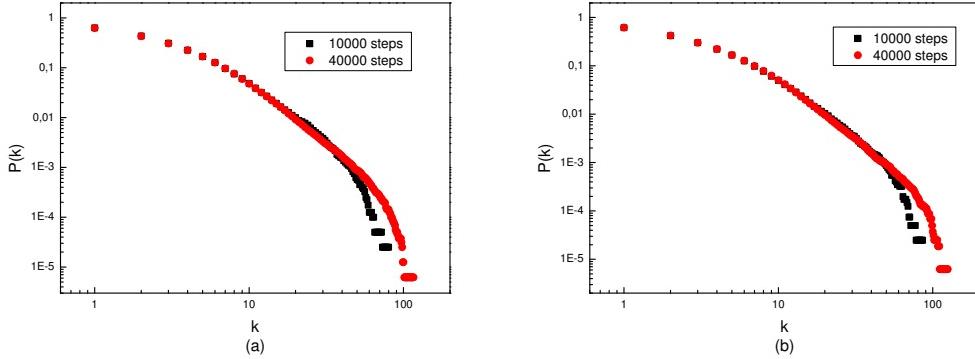


Figure 7: (a) Degree distribution of authors in the Model 1 with different system size 10000 steps and 40000 steps. (b) Degree distribution of authors in the Model 2 with different system size 10000 steps and 40000 steps.

\max_n , \max_m and the decreasing of λ , the degree distribution of authors is far from a pow law distribution in theses two models. Fig. 8 and 9 show the cumulative degree distribution of authors and papers obtained by Model 1 and Model 2 respectively. The simulations are consistent with the scale free property observed from empirical data. To check the validity of our two models, we also proceed to compare the empirical data on *Econophysicist*, *Complex – networks*, *Cooccurrence*, *Coauthoring* networks with numerical simulations of two models made above (see Fig. 10).

4 Conclusion

Many real-world networks display a natural bipartite structure. This makes us study them should be as bipartite ones. In this paper, we collect some representation real-world bipartite networks and according to the different relation of two sets of nodes, divide them into two types, dependence bipartite networks and independence bipartite networks. In dependence networks, one kind of nodes is the results or groups which are gotten by some kind of behaviors among the other kind of nodes. The independence network is opposite with the dependence network, two sets exist independently. By analyzing the degree distribution of these collected networks, we find the actors nodes have scale-free property in the dependence networks, and there is no accordant degree distribution in the independence networks for both two types of nodes.

Noticing the scale-free property of actors in dependence networks, we propose two models for generation of dependence bipartite graphs. This two models are both growing models, but are not based on the preferential attachment mechanism. The obtained networks fit well the properties of empirical results, using only their general bipartite structure. However, there still have other properties not obtained by the bipartite

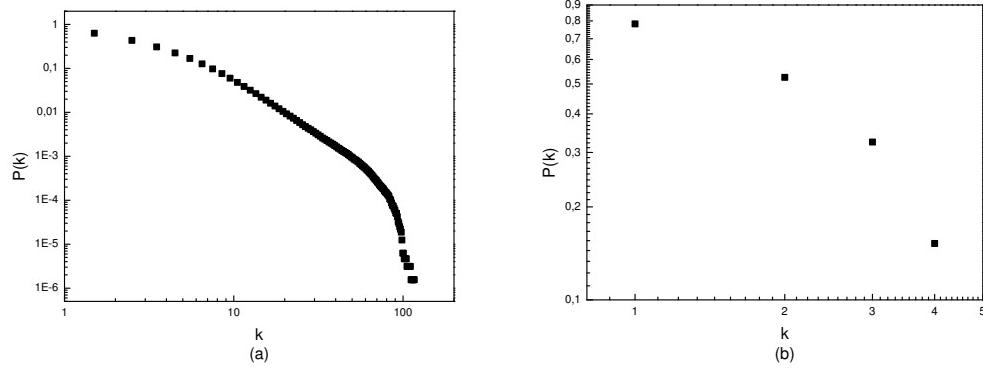


Figure 8: (a) Degree distribution of authors gotten by Model 1 with parameters $\max_n=5$, $\lambda=0.8$, $\max_m=2$. It obeys a pow law well. (b) Degree distribution of papers gotten by Model 1 in the same parameters.

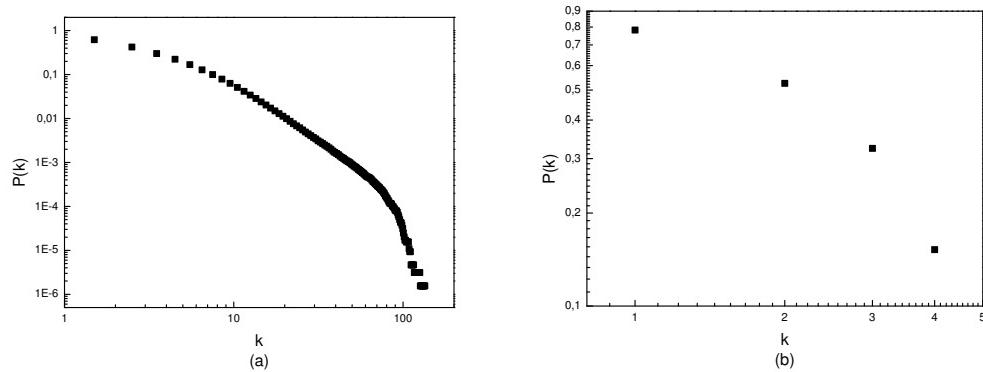


Figure 9: (a) Degree distribution of authors gotten by Model 2 with parameters $\max_n=5$, $\lambda=0.8$, $\max_m=2$. It also obeys a pow law well. (b) Degree distribution of papers gotten by Model 2 in the same parameters.

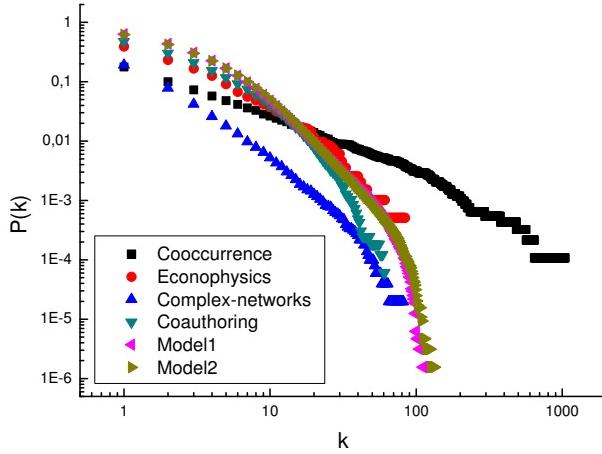


Figure 10: Degree distribution of actors, including empirical results from *Econophysicist*, *Complex – networks*, *Cooccurrence*, *Coauthoring* and simulation results from *Model1*, *Model2*.

models, The independence network is not studied well. There are a lot of works to do in bipartite networks.

5 Acknowledgement

The author P. Zhang thanks Matthieu Latapy for sharing the network data. This work is partially supported by 985 Project and NSFC under the grant No. 70771011 and No. 60534080.

References

- [1] R. Albert, A.-L. Barabási, Rev. Mod. Phys. 74 (2002) 47.
- [2] M.E.J. Newman, SIAM Rev 45 (2003) 167.
- [3] S. N. Dorogovtsev, J. F. F. Mendes, Adv. Phys. 51, 1079 (2002).
- [4] R. Albert, H. Jeong, A.-L. Barabási, Nature 401 (1999) 130.
- [5] Dorogovtsev S N, Mendes J F F, Evolution of Networks: From Biological Nets to the Internet and WWW, Oxford University Press, 2003.
- [6] M.E.J. Newman, Proc. Natl. Acad. Sci. 101 (2004) 5200-5205.

- [7] P. Gleiser, L. Danon, *Advances in Complex Systems*, 6 (2003) 565-573.
- [8] R. J. Williams, N. D. Martinez, *Nature* 404 (2000) 180-183.
- [9] Joshua R. Tyler, Dennis M. Wilkinson, Bernardo A. Huberman, arXiv:cond-mat/0303264.
- [10] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, A. Arenas, *Phys.Rev. E* 68 (2003) 065103.
- [11] Gergely Palla, Imre Derényi, Illés Farkas, Tamás Vicsek, *Nature* 435 (2005) 814.
- [12] D. Watts, S. Strogatz, *Nature* 393 (1998) 440.
- [13] Michel L. Goldstein, Steven A. Morris, and Gary G. Yen, *Phys.Rev. E* 71 026108 (2005).
- [14] M.E.J. Newman, *Phys.Rev. E* 64 016131 (2001).
- [15] M.E.J. Newman, *Phys.Rev. E* 64 016132 (2001).
- [16] Borgatti, S.P., Everett, M.G., Network analysis of 2-mode data. *Social Networks* 19 (3), 243269, 1997.
- [17] Matthieu Latapy, Clémence Magnien, Nathalie Del Vecchio, *Social Networks*, 30 (1), 31-48, Jan 2008.
- [18] Liljeros, F., C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, 2001, *Nature* (London) 411, 907.
- [19] Guillaume,J.-L., Le Blond, S., Latapy, M., Statistical analysis of a P2P query graph based on degrees and their time-evolution. In: Proceedings of the 6th InternationalWorkshop on Distributed Computing (IWDC), Lecture Notes in Computer Sciences (LNCS), 2004.
- [20] Guillaume, J.-L., Le Blond, S., Latapy, M., Clustering in P2P exchanges and consequences on performances. In: Proceedings of the 4th International Workshop on Peer-to-Peer Systems (IPTPS), Lecture Notes in Computer Sciences (LNCS) 2005.
- [21] J. J. Ramasco, S. N. Dorogovtsev, and R. Pastor-Satorras, *Phys. Rev. E* 70, 036106 (2004).
- [22] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Proc. Natl. Acad. Sci. U.S.A.* 99, 2566-2572 (2002).
- [23] Jean-Loup Guillaume, Matthieu Latapy, *Information Processing Letters* 90 (2004) 215221.
- [24] Jean-Loup Guillaume, Matthieu Latapy, *Physica A* 371 (2006) 795813.

- [25] A. Ramezanpour, *Europhys. Lett.* 68, 316 (2004).
- [26] R. Lambiotte, M. Ausloos, *Phys.Rev. E* 72 066107 (2005).
- [27] A.-L. Barabási and R. Albert, *Science* 286, 509 (1999).
- [28] Jun Ohkubo, Kazuyuki Tanaka, Tsuyoshi Horiguchi, *Phys. Rev. E* 72, 036120 (2005).
- [29] Ying Fan, Menghui Li, *et al*, *International Journal of Modern Physics B* Vol. 18, Nos. 17-19 (2004) 2505-2511.
- [30] R. Ferrer, R.V. Solé, The small-world of human language, *Proceedings of the Royal Society of London B* 268 (2001) 22612265.